The Smithsonian/NASA Astrophysics Data System Status Report

Alberto Accomazzi Michael Kurtz Harvard-Smithsonian Center for Astrophysics http://ads.harvard.edu







The ADS Project

- Established in 1989 (before the web!) as a portal for accessing astronomical data and bibliographic metadata
- Was restructured in 1994 to become an A&I service for astronomers and astrophysicists, with fulltext archive
- Has 100% penetration in astronomical community, with take-up in other areas of space sciences, engineering and physics

History of the ADS project (1/2)

- 1992: literature database implemented
- 1993: implemented federated search to object database
- 1994: web-based version of ADS launched
- 1995: fulltext of ApJ Letters digitized and online
- 1996: citation data incorporated in ADS; links between bib records and datasets created; first mirror site online
- 1997: indexed astronomy preprints from arXiv
- 1998: online readership via ADS > worldwide print readership
- 1999: incorporated 1.2M citations via text mining

History of the ADS project (2/2)

- 2000: incorporated usage data, citation ranking
- 2001: 10th mirror site comes online
- 2002: 300K historical scans from microfilms online
- 2003: myADS notification service launched
- 2004: introduced fulltext search, private libraries
- 2005: introduced daily database updates, RSS feeds
- 2006: introduced basic search, openURL links
- 2007: implemented user login system
- 2008: introduced topic search

ADS Collaborators



Who uses ADS

- Astronomers & Astrophysicists
- Solar physicists, planetary scientists
- Space scientists, engineers
- Geophysicists
- Physicists
- Amateur astronomers

ADS Data Holdings

- Over 8M bibliographic metadata records:
 - Astronomy: I.7M
 - Physics: 5.4M
 - Arxiv e-prints: 590K
 - Citations: 40M (over 3.4M papers with citations)
 - Curated links: 23M (fulltext, data products, citations)
- Fulltext archive of literature:
 - 4M scanned pages, 544K articles
 - 650K pages historical material

Over past 12 months

- ingested 800K metadata records
- bibliographic metadata coverage up by 10%
- citations up by 20% to 40M
- ingested 336K pages of fulltext (180K pages historical)

ADS Usage

- Over 40K registered users (with login account)
- Over 8,000 myADS users signed up for notifications
- Over 30K regular users (> 10 queries/month)
- I.2M users from search engines every month
- 8M database queries every month
- 16M bibliographic records retrieved every month
- 2.5M scanned article pages downloaded every month
- I.2M fulltext downloads via ADS every month

Operations

- Collaborations and agreements with over 100 partners:
 - 21 major publishers contributing ~600 journals
 - 72 minor publishers contributing 100+ journals
 - CrossRef metadata services
 - astrophysics data archives, observatories, missions
 - other "like-minded" projects

Metadata Ingest

- Coverage: Astronomy, Planetary Sciences Physics, Geophysics, general science (Nature, Science, PNAS, etc.)
- Journals, conference proceedings, reports, observing proposals, bulletins, newsletters, PhD Thesis
- Database updates: daily from e-prints, weekly for Astronomy database, 2x month for Physics database
- Metadata feeds, completeness, quality all over the map; lots of effort goes in curation, merging of metadata, tracking provenance:
 - high-throughput, high/low quality (publishers)
 - low-throughput, hand-generated (observatories, other projects, small publications), conference proceedings, staff and contractors

Future Plans

- Restructure system architecture
- Improve database completeness and metadata curation
- Improve user interface
- Extend search capabilities
- Enhance personalization, recommendations
- Incorporate semantic web technologies

System Architecture

- New architecture for metadata curation and search based on CERN Invenio/INSPIRE
- Currently extending Invenio search interface and capability to match ADS's requirements
- Implementing pipeline for metadata ingest in new system from ADS "classic"
- Goal: alpha system online by Jan 2011

System Customizations

- ADS-style author indexing and searching
- ADS-style word parsing and indexing
- Implementation of synonym search from knowledge base (authors, text words)
- Simplified use of word weights for scoring
- Federated search of object databases
- Selection/filtering by bibliographic groups

ADS Labs

- A playground for testing new ideas, technologies
- A bridge present with the future, without having to worry about legacy too much
- Aim: receive feedback from users, test performance and impact of new tools, services
- It's OK if we break things in the labs

Labs Ideas

- Metadata enhancement (normalized keywords, affiliations, ORCID experiments)
- Infrastructure (caching, federated searching, RDF output)
- New services (recommendation, fulltext search)
- User interface (search, facets, widgets, topic browser)

New User Interface

- Looks TBD
- Remove clutter, hide complexity, use javascript/AJAX where appropriate
- Use facets throughout
- Provide recommendations, context when appropriate

Fulltext Search

- Enable searching and viewing of papers in ADS fulltext archive
- Allow searching of current content, with links to publisher's fulltext offerings
- Support text mining efforts, metadata enrichment
- Currently investigating available technologies (SOLR, Invenio)

Links to Data

- ADS provides links to data catalogs, astronomical objects, data archives
- Metadata about links provided by publishers, collaborators, librarians
- Involved in VAO Data Curation & Preservation effort to expand data linking via semantics
- Goal: enable data discovery, use literature as filter on data (and vice-versa)

Notifications: myADS

Collaborators							
		myADS Notification for Ast	ronomy				
	Apple Xabob Coor	narvard.edu/myADS/cache/268396607_AST.html	C Google				
	La III Apple Tanoo dogle Maps Tourube Wikipedia News (263) * Popular*						
	mil.	myADS Personal Notification Service for Alberto Accomazzi Tue Nov 10 04:57:53 2009 Astronomy database					
	ads						
	NIACO						
My citations —	ADS Main Queries	ACCOMAZZI, ALBERTO - Citations: 223 (total 242)	Favorite Authors - Recent Papers				
	Astronomy	2010ARIST.443K: Kurtz,+: Usage Bibliometrics	2009Ap&SS.322101R: Rosen,+: Laboratory experiments				
	Physics arXiv e prints	near-infrared core source in 3C 433	clumpy environments				
	FAO	2009MNRAS.399.1622G: Goyal,+: Unusual optical					
	What's new	intranight time-scale	(ASTROPHYSICS DATA SYSTEM, etc				
		2009ASPC.411.384H: Henneken,+: Exploring the	2009amos.confE87S; Schmalz.+: Comparison of Neural				
	Current Tables of Contents	2009IISys2474C: Chen,+: Mapping the Sloan Digital Sky	Networks and Tabular Nearest Neighbor Encoding for				
	Astronomical Journal	Survey's Global Impact	Hyperspectral Signature Classification in Unresolved Object Detection				
Pocont papars	Astrophysical Journa	"VIDTUAL ORSEDVATODV" at a Bagant Banars	2009MNRAS.399.2231S: Scaringi,+: Classifying broad				
Recent papers	Astrophysical Journal Letters	2009Obs129397P: Prema: Galaxy formation and	absorption line quasars: metrics, issues and a new catalogue				
	Monthly Notices of the Royal	evolution using the Virtual Observatory					
	Publications of the	2009LNP79181W: Walton,+: AstroGrid and the Virtual Observatory	(ASTROPHYSICS DATA SYSTEM, etc -				
	Astronomical Society of	2009A&A506455S: Solano,+: The LAEX and NASA	Most Popular 2008 ApJ 674, 7680: Ovaizu +: A Galaxy Photometric				
	Japan Bacant Astro PH	2009MNRAS 399.2231S: Scaringi.+: Classifying broad	Redshift Catalog for the Sloan Digital Sky Survey Data				
	Kecchi ASUO-PH	absorption line quasars: metrics, issues and a new catalogue	Release 6				
	Search Recent Papers	constructed from SDSS DR5 2009SoPhtmp.,155U; Ulrich.+: Modeling Total Solar	segregation using spectral and integrated band data for				
	Author	Irradiance Variations Using Automated Classification	TAUVEX/ASTROSAT satellite data pipeline				
	Subject	Software on Mount Wilson Data	luminosity functions in the Sloan Digital Sky Survey				
Hottest papers	Send Query	"VIRTUAL OBSERVATORY", etc Most Popular	2009MNRAS.392.233Z: Zhang,+: Morphology				
i ioccese papers	Update myADS Settings	2009ASPC411101S: Seaman,+: Optimal DN Encoding	galaxies				
			- (7/)				



A.Accomazzi - The Smithsonian/NASA Astrophysics Data System

IWGDD 11/19/2009

ADS Query Results Image: ADS Query										
Query Results from the	e Maps YouTube Wikipedia News (2 s Data System (ADS) ADS Database	objects in papers	Go to bottom of page							
Related Objects	Selected and retrieved 200 abstracts	Sort options								
NAME HDF (12) NAME Chandra Deep Field- South (8) ACO 1689 (8) ACO 901 (8) ZwCl 0024+1652 (7) ACO 902 (7) ACO 2390 (6) ZwCl 1455+2232 (5) NAME SGP (4) CIG 2137.3-2353 (4) ACO 2261 (4) ACO 2219 (4) ACO 1763 (4) ACO 209 (4) ACO 68 (4) 1E 0302.5+1717 (4) ZwCI 1358+6245 (3) NAME HDF-S (3) NAME COSMOS FIELD (3) NAME ACO 901A (3) NAME ACO 901-902	# Bibcode Authors	Score DateList of LinksTitleAccess Control Help								
	1 <u>□ 2006glsw.book269S</u> Schneider, P.	81.000 n/a 2006 A E X R C Weak Gravitational Lensing	<u>c U</u>							
	2	68.000 Jun 2008 <u>A E X</u> <u>R C</u> Cosmology with weak lensing surveys	<u>c U</u>							
	3 <u>2003ARA&A41645R</u> Refregier, Alexandre	61.000 n/a 2003 A E F X R C Weak Gravitational Lensing by Large-Scale Structure	<u>c</u> <u>UH</u>							
	4	51.000 Nov 2008 A X R C Weak Gravitational Lensing and Its Cosmological Applica	<u>c</u> <u>U</u> ations							
	5 <u>2003astro.ph6465S</u> Schneider, Peter	44.000 Jun 2003 A X R C Gravitational lensing as a probe of structure	<u>с <u>U</u> <u>H</u></u>							
SUPERCLUSTER (3) CIG 1224.7+2007 (3) CIG 1054.4-0321 (3) ACO 1942 (3)	6	41.000 May 2006 <u>A</u> <u>E</u> <u>G</u> <u>X</u> <u>R</u> <u>C</u> The Shear Testing Programme - I. Weak lensing analysis observations	c U of simulated ground-based							

A.Accomazzi - The Smithsonian/NASA Astrophysics Data System

IWGDD 11/19/2009

A.Accomazzi - The Smithsonian/NASA Astrophysics Data System

IWGDD 11/19/2009

Recommendations

ads	The S	mithsonian/I	NASA Astrophy	sics D	ata Sys	stem	troud Supporter of
Home	Help	Sitemap				2005MNRAS.359308Z	Search
Free Fulltext A	rticle	Citations	Find Similar Articles	Full	record info		
 Accreting black holes Życki, Piotr T.; Niedźwiecki, Andrzej Monthly Notices of the Royal Astronomical Society, Volume 359, Issue 1, pp. 308-314. X-rays produced by compact flares corotating with a Keplerian accretion disc are modulated in time by Doppler effects. We improve on previous calculations of these effects by considering recent models of intrinsic X-ray variability, and we compute the expected strength of the relativistic signal in current data of Seyfert galaxies and black hole binaries. Such signals can clearly be seen in, for example, recent XMM-Newton data from MCG-6-30-15, if indeed the X-rays are produced by corotating flares concentrated toward the inner disc edge around an extreme Kerr black hole. The lack of the signal in the data collected so far gives support to models where the X-ray sources in active galaxies do not follow Keplerian orbits close to the black hole. Keywords: accretion, accretion discs - relativity - galaxies: active - X-rays: binaries - X-rays: individual: MCG-6-30-15. DOI: 10.1111/j.1365-2966.2005.08887.x 				sc are of we eyfert s are ind an gives	 2008ApJ679L37L Liu et al, Precise Measurement of the Spin Parameter of the Stellar-Mass Black Hole M33 X-7 2006ApJ646394M Miller et al, Simultaneous Chandra and RXTE Spectroscopy of the Microquasar H1743-322: Clues to Disk Wind and Jet Formation from a Variable Ionized Outflow 2009NewA14674F Foellmi, What is the closest black hole to the Sun? 2009ApJ695888U Ueda et al, GRS 1915+105 in "Soft State": Nature of Accretion Disk Wind and Origin of X-ray Emission 1973A&A24337S Shakura et al, Black holes in binary systems. Observational appearance. 2007ARA&A45441M Miller, Relativistic X-Ray Lines from the Inner Accretion Disks Around Black Holes 2006Apt44730M McHardy et al, Active galactic nuclei as scaled-up Galactic black holes 		

The ADS is Operated by the Smithsonian Astrophysical Observatory under NASA Grant NNX09AB39G

E RAPA

Conclusions

- ADS project has 100% penetration in community, used by other disciplines (physics) and general public
- Current project focus is IT restructuring, technology realignment, UI redesign (2 yr timeline)
- Many new features and technologies being tested and considered, ready to collaborate with partners (e.g. ORCID)
- Future plans include support for text mining, linked data applications, integration of observational metadata, increased personalization & recommendations

Demo

- Faceted Topic Search
- <u>Keyword Topic Cluster</u>
- Article Recommeder (alpha)